International Journal of Future Engineering Innovations

Industrial Knowledge Graph Framework for Autonomous Root Cause Analysis and Predictive Process Optimization

Hande Sonal [1*], Waman Shruti [2], Bankar Swarupa [3], Haral Shrushti [4], Dr. Khatal SS [5]
[1-5] Department of Computer Engineering, Sharadchandra Pawar College of Engineering, Savitribai Phule Pune University, Pune, Maharashtra 412409, India

* Corresponding Author: **Hande Sonal**

Article Info

ISSN (online): 3049-1215
Volume: 03
Issue: 06
Received: 06-09-2026
Accepted: 07-10-2026
Published: 05-11-2026
Page No: 76-81

Abstract
Finding the reason behind a machine fault is not always simple. Industrial equipment produces sensor readings throughout the working day, and a change in one part of a machine may affect other parts as well. This paper presents an Industrial Knowledge Graph Framework intended to bring machine-learning methods and connected equipment information into one diagnostic process. Isolation Forest is used to flag readings that look unusual, and XGBoost estimates the risk of equipment failure. The Knowledge Graph records how machines, components, sensors, faults, and maintenance activities relate to one another. By following these links, maintenance staff can investigate possible causes and consider how a fault might spread. The framework is also intended to suggest maintenance actions and show the findings on a dashboard. The overall aim is to give staff earlier warning of developing problems and useful context for deciding what to inspect.

Keywords: Industrial Knowledge Graph, Root Cause Analysis, Anomaly Detection, Failure Prediction, Predictive Maintenance, Fault Propagation, Machine Learning

1. Introduction

Machines, sensors, and automated controls are closely linked in most industrial production settings. While equipment is running, sensors record values such as temperature, pressure, vibration, rotational speed, and machine status. These measurements help operators follow the condition of a machine, but they do not necessarily tell them why a fault has appeared. The difficulty increases when one component affects another or when several process variables change at the same time. A fault that is found late may interrupt production, increase repair costs, and reduce operating efficiency [1, 2].

Machine-learning methods can help engineers examine these readings, recognize behaviour that differs from the usual pattern, and estimate the risk of a future failure. Still, a

warning from a model is only one part of the diagnosis. It may identify an unusual reading without showing what caused it or which connected parts should also be inspected. Equipment history and component relationships are therefore important when interpreting an alert. Effective AI-driven diagnostic systems should support human interpretation rather than rely exclusively on automated predictions. Tasleem *et al*. (2024) proposed a decision intelligence framework that integrates AI-generated predictions with human judgment, contextual evaluation, and iterative feedback. This approach provides a relevant conceptual foundation for the proposed industrial knowledge graph framework, in which automated fault detection and failure-risk estimates support maintenance personnel in identifying potential causes and making informed corrective decisions.

A Knowledge Graph provides a practical way to keep those relationships together. It can connect each machine with its components and sensors, and link these items to known faults and maintenance records. Earlier work on knowledge-based diagnosis and graph-based anomaly detection indicates that such links can assist fault investigation [3-5]. In the proposed framework, Isolation Forest highlights unusual operating conditions and XGBoost estimates failure risk [11]. The detected event is then associated with relevant graph entities. This allows the system to explore likely causes and possible paths through which a fault could affect other equipment [12].

The framework is intended to give maintenance personnel evidence they can examine, rather than presenting an alert without context [19]. Its results will depend on several practical factors: whether the input records are complete, whether the prediction models are reliable, and whether the Knowledge Graph reflects the actual equipment connections [16-18].

2. Problem Statement

The central problem addressed in this project is that an abnormal reading does not always reveal the origin or seriousness of a machine fault. Industrial equipment generates large amounts of operating data, and the links between machines, components, and process variables make diagnosis difficult. Conventional monitoring can flag a condition that falls outside its normal range, but it may not show where the fault began, how it could affect other equipment, or whether a failure is likely. If staff do not have this context, maintenance may be delayed and unplanned stoppages and costs may increase. The proposed framework combines anomaly detection, failure-risk estimation, Knowledge Graph reasoning, and maintenance guidance to help address these gaps.

3. Objectives

- To collect and preprocess industrial sensor data for reliable machine condition analysis.
- To detect abnormal machine behavior using the Isolation Forest algorithm.
- To predict equipment failure risk using the XGBoost algorithm.
- To construct a Knowledge Graph for identifying probable root causes and tracing fault propagation.

4. Literature Survey

1. Cai *et al*. (2024) presented ―Research on Equipment Fault Diagnosis Method Based on Knowledge Graph in Intelligent Manufacturing‖. They construct a multi-level Knowledge Graph to organize information collected from multiple sources for diagnosing defects in rotating machinery. We employ Bayesian analysis and reasoning about relational routes to reduce the set of potential fault origins. The results indicate how linkages between equipment can contribute to diagnosis. This approach may be integrated with a failure risk estimation model, and recommendations to assist staff in choosing the action to take, to offer a broader maintenance solution.
2. Xia *et al*. (2023) named ―Recommendation of Maintenance Planning for Complex Industrial Equipment Based on Knowledge Graph and Graph Neural Network.‖ They use a manufacturing Knowledge Graph and an attention-based compressed relational graph convolutional network to associate defects with potential maintenance solutions. It also has a question-and-answer option to fetch pertinent information. That could assist maintenance personnel discover guidance without having to dig through records. Future studies could integrate anomaly detection, failure-risk prediction, and fault propagation analysis in one method.
3. Mao *et al*. (2024) A Knowledge Graph keeps equipment information connected. It can link machines with their components and sensors, then connect these items with known faults and maintenance records. Earlier research on knowledge-based diagnosis and graph-based anomaly detection indicates that these relationships can support fault investigation [3-5]. In the proposed framework, Isolation Forest identifies unusual operating conditions and XGBoost estimates failure risk [11]. The detected event is linked to relevant graph entities so that likely causes and possible fault paths can be explored [12].
4. Wang *et al*. (2024) reviewed ―Knowledge Graph Based In-Context Learning for Advanced Fault Diagnosis in Sensor Networks‖. Their solution integrates input from sensor networks with expert knowledge in a domain-specific Knowledge Graph. The knowledge recorded about sensor conditions can aid in fault identification and isolation. This study is applicable here as a reading is easier to interpret when seen with information about the equipment. Extensions to this approach are conceivable to predict failure risk and provide maintenance measures.
5. Kim and Lee (2026) proposed ―Maintenance Knowledge Graph and Time-Series Relational Graph Convolution Neural Networks for Enhancing Predictive Maintenance of Wind Turbine Gearbox.‖ The framework joins a maintenance Knowledge Graph with a time-based relational graph convolutional network. It examines equipment degradation and supports fault classification and maintenance-action prediction. This is closely related to the present project, although further evaluation would be needed to establish how well the method transfers to other industrial machines and broader root-cause investigations.
6. Wen *et al*. (2026) presented ―KG-EIGNN: A Knowledge Graph Based Graph Neural Network for Fault Detection in Industrial Systems‖. Their model uses sensor data, expert knowledge and spatiotemporal correlations to improve fault detection. Knowledge-guided message passing, supervised learning, and contrastive learning help address imbalanced data and support generalization. The work illustrates the potential of graph-based fault detection. A practical next step would be to combine this type of diagnosis with a simpler anomaly detector, failure-risk estimates, and tools that help maintenance teams choose their next action.

Table 1: Comparison Table

Author & Year	Method Used	Advantages	Limitations
Cai *et al*. (2024)	Knowledge Graph, Bayesian reasoning	Identifies fault causes	Focuses on equipment diagnosis
Xia *et al*. (2023)	Knowledge Graph, GNN	Supports maintenance planning	Limited integrated analysis
Mao *et al*. (2024)	Graph Convolutional Network	Enables fault reasoning	Limited maintenance integration
Wang *et al*. (2024)	Knowledge Graph-based learning	Improves fault identification	Depends on domain knowledge
Kim and Lee (2026)	Temporal Knowledge Graph, R-GCN	Supports predictive maintenance	Requires broader validation
Wen *et al*. (2026)	Knowledge Graph, GNN	Supports fault detection	Limited decision-support integration

5. Working of System

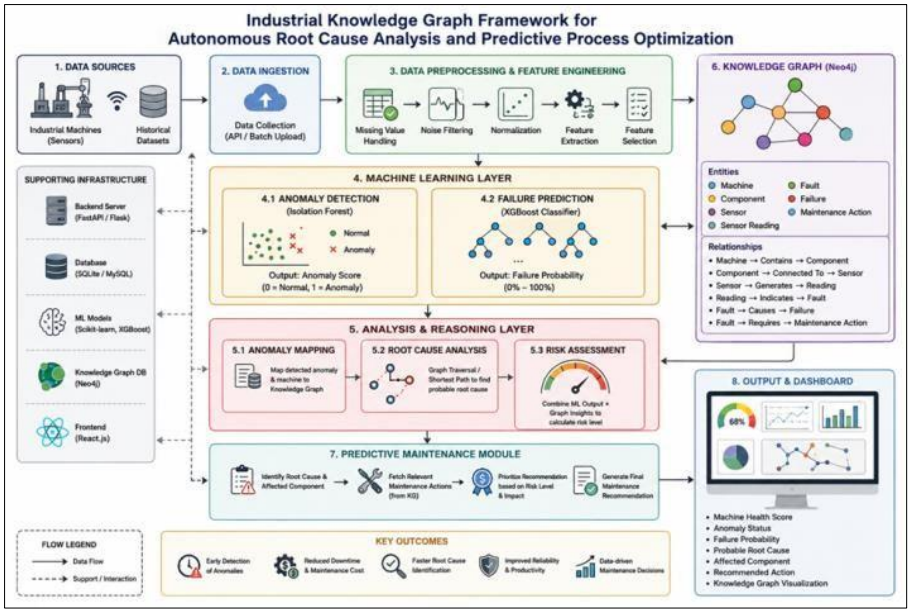


Fig 1: Design of the system

6. System Design

6.1. Data Sources

The objective of the framework is to provide maintenance people evidence to investigate, instead of an alert without context [19]. The usefulness of this solution will depend on the quality of the accessible records, the performance of the prediction models, and the representation of the Knowledge Graph to the actual equipment and its connections [16-18].

6.2. Data Ingestion

A common data pipeline is fed with information from diverse sources. This can be fed into the system using application programming interfaces or staff can batch upload older records. The data-ingestion step organizes the incoming information into a standard format that the subsequent processing steps may use.

6.3. Data Preprocessing and Feature Engineering

The raw records are often processed before they can be analyzed. This stage checks for missing data, removes unnecessary noise and scales numeric fields if necessary. It also gets the records into a uniform format. The system extracts important elements from observations like temperature, vibration, pressure and rotational speed and then preserves only the most relevant variables to the analysis.

6.4. Machine Learning Layer

It was a quiet evening, sky all striped with pink and purple. I walked down the vacant path to the soft singing of birds. There was a kind of beauty to this moment, like time decided to take a little breather. I stopped, took a breath and just let the serene beauty of nature settle in. Funny how it's the little moments that bring you the most peace.

6.4.1. Anomaly Detection

An industrial production system consists of machinery, sensors and automated controls. In operation, sensors make measures including temperature, pressure, vibration, rotational speed and machine condition. These readings assist operators monitor equipment, although a single reading rarely reveals the cause of a malfunction. The
problem with the research grows when a component influences another or a number of readings alter together. Unnoticed defects can result in production stoppages, increased maintenance costs, and reduced operating efficiency [1, 2].

6.4.2. Failure Prediction

The problem addressed here is that an abnormal sensor reading does not necessarily reveal the source or seriousness of a machine fault. Industrial equipment generates large volumes of data, and the connections among machines, components, and process variables make diagnosis challenging. Existing monitoring tools may flag a value outside its normal range but may not show where the fault began, what other equipment could be affected, or whether a failure is likely.

6.5. Analysis and Reasoning Layer

At this stage, the system considers three sources together: anomaly alerts, the failure-risk estimate, and the relationships recorded in the Knowledge Graph. The combined information helps investigate possible explanations for unusual behaviour. It can also indicate which connected components may be affected and show possible routes along which a fault could travel.

6.5.1. Anomaly Mapping

After an unusual condition is flagged, the anomaly-mapping step connects it to the appropriate entities in the Knowledge Graph. The available records may link the event to a particular machine, sensor, component, or previously known fault. This gives the next stage more context for investigating the alert.

6.5.2. Root Cause Analysis

To investigate the source of a problem, the root-cause module follows links in the Knowledge Graph among machines,

components, sensors, readings, and known fault conditions. The relationships can point to likely explanations that deserve inspection. They should be treated as possible causes until maintenance staff verify them against the actual equipment.

6.5.3. Risk Assessment

The risk-assessment step combines the model's failure estimate with the findings from root-cause analysis. It considers the fault type, the components that may be involved, and the operating information available. Beyond generating predictive estimates, AI-driven risk assessment systems must account for uncertainty when supporting decisions. Tasleem *et al*. (2026) demonstrated how Bayesian machine learning can incorporate uncertainty estimates into decision intelligence frameworks. Although developed for technology readiness and international investment, their approach highlights the broader importance of uncertainty-aware decision support, which is conceptually relevant to interpreting equipment failure-risk predictions alongside diagnostic evidence.

Machine-learning methods can help engineers review operating data, identify patterns that differ from normal behaviour, and estimate the possibility of failure. However, a model alert does not always explain why a reading changed or which connected components may need attention. For a useful diagnosis, the alert should be considered together with the equipment's history and the relationships between its parts**.**

The Knowledge Graph represents industrial information as connected items rather than separate records. Items may describe machines, components, sensors, sensor readings, faults, failures, and maintenance activities.

6.6. Predictive Maintenance Module

The predicted risk of failure, probable cause, and data on the components involved are used to produce maintenance recommendations. The next step could be an inspection, repair or scheduled maintenance activity, depending on the records available and the technical knowledge possessed. Staff can see the proposal and decide what to do next.

6.7. Supporting Infrastructure

The framework needs several software components to be run. The application logic is hosted on a backend server, which prepares the data, runs the models and arranges communication between modules. Equipment information, sensor readings, anomaly results, failure estimates and maintenance advice are stored in a database. The Knowledge Graph database contains the entities and the links between them.

6.8. Overall System Workflow

The records are checked, prepared, and converted into useful features. Isolation Forest checks for unusual patterns, while XGBoost estimates failure risk. Any flagged event is linked to related items in the Knowledge Graph, where the connections can be examined for likely causes and possible fault-propagation routes. The combined results can then inform maintenance planning.

7. Expected Outcome

The planned framework brings monitoring, fault investigation, and maintenance planning into a single process. While XGBoost predicts the likelihood of equipment failure, the Isolation Forest identifies abnormal operating conditions. The Knowledge Graph links equipment, components, sensors, faults, failures and maintenance operations, enabling personnel to explore potential reasons and how a fault may propagate. These results can be used to provide recommendations for maintenance. The expected outputs include earlier alerts, a clearer view of the links between faults and equipment, and a dashboard for reviewing results. As the framework is proposed rather than validated, its performance still needs to be measured using appropriate industrial data.

Table 2: Expected Outcomes of the Proposed System

Sr. No	Expected Outcome
1	Early anomaly detection using Isolation Forest.
2	Failure-risk prediction using XGBoost.
3	Knowledge Graph construction to represent industrial entities and their relationships.
4	Probable root cause identification and fault propagation analysis.
5	Predictive maintenance recommendations based on identified faults and predicted risks**.**
6	Interactive dashboard for industrial monitoring and decision support.

8. Conclusion

The project proposes an Industrial Knowledge Graph Framework for using Machine Learning and connected equipment information together during fault diagnosis and maintenance planning. Isolation Forest flags unusual operating patterns, and XGBoost estimates possible equipment failure risk. The Knowledge Graph records links among equipment, components, sensors, faults, and maintenance activities, which can help investigators examine likely causes and possible fault paths. The findings may support decisions about inspection and repair. Since this is a proposed framework, its effectiveness cannot be assumed; it must be tested with suitable industrial data and reviewed by people familiar with the equipment.

9. Future Scope

Future work can be developed to enhance the suggested Industrial knowledge graph framework by incorporating real-time sensor data from the IoT for continuous tracking of the equipment and more rapid detection of defects. We may also employ advanced deep learning and graph neural network techniques to improve failure prediction, root cause identification and fault propagation analysis. The knowledge graph can be updated constantly with fresh equipment data, maintenance logs, and changing failure patterns. Furthermore, reinforcement learning can be applied to optimize maintenance schedules and resource allocation. Integration of industrial digital twins and cloud-based systems to enable wide scale monitoring of several manufacturing units.

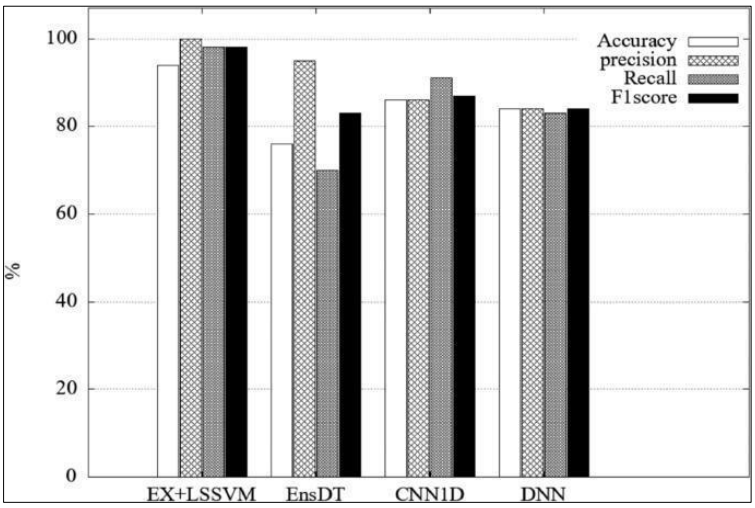


Fig 2: Expected Model Evaluation Metrics Different measures will be used for different parts of the framework. The evaluation is intended to check prediction quality, retrieval of similar cases, the usefulness of summaries, how well predictions can be explained, and the consistency of data handling and processing.

Table 3

Component	Evaluation Metrics
Outcome Prediction	Accuracy, Precision, Recall, F1- score
Similar Case Retrieval	Precision@K, Recall@K
Case Summarization	ROUGE scores and human evaluation
Prediction Explainability	Feature-importance analysis and expert review
System Performance	Response time and processing time
Case Management	Successful operations and data consistency

References

1. Cai C, Jiang Z, Wu H, Wang J, Liu J, *et al*. Research on knowledge graph-driven equipment fault diagnosis method for intelligent manufacturing. Technology. 2024;130:4649-4662. doi:10.1007/s00170-024-12998-x Springer Nature Link
2. Wen, *et al*. introduced —KG-EIGNN: A Knowledge Graph Based Graph Neural Network for Fault Detection in Industrial Systems.‖ ResearchGate. 2026. doi:10.1109/ACCESS.2025.3558631
3. Liu FT, Ting KM, Zhou ZH. Isolation Forest. In: Proceedings of the IEEE International Conference on Data Mining, 413–422. DOI; 2008. doi:10.1109/ICDM.2008.17
4. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. Xia *et al*. (2023) described — Maintenance Planning Recommendation of; 2016. doi:10.1145/2939672.29397852
5. Bo W, Ju C, Hu J, *et al*. KG-EIGNN: A knowledge graph based graph neural network for fault detection in industrial systems. Scientific Reports. 2026;16:15471. doi:10.1038/s41598-026-42723-5
6. Hogan A, Blomqvist E, Cochez M, *et al*. Knowledge graphs. ACM Computing Surveys. 2021;54(4):71. doi:10.1145/3447772
7. Ji S, Pan S, Cambria E, Marttinen P, Yu PS. A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems. 2022;33(2):494-514. doi:10.1109/TNNLS.2021.3070843
8. Nickel M, Murphy K, Tresp V, Gabrilovich E. A review of relational machine learning for knowledge graphs. Proceedings of the IEEE. 2016;104(1):11-33. doi:10.1109/JPROC.2015.2483592
9. Wang, Q., Mao, Z., Wang, B., & Guo, L. (2017).
10. Knowledge graph embedding: A survey of approaches and applications. IEEE Transactions on Knowledge and Data Engineering, 29(12), 2724–2743. https://doi.org/10.1109/TKDE.2017.2754499.
11. Breiman L. Random forests. Machine Learning. 2001;45:5-32. doi:10.1023/A:1010933404324
12. Friedman JH. Greedy function approximation: A gradient boosting machine. The Annals of Statistics. 2001;29(5):1189-1232. doi:10.1214/aos/1013203451
13. Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey. ACM Computing Surveys. 2009;41(3):15. doi:10.1145/1541880.1541882
14. Susto GA, Schirru A, Pampuri S, McLoone S, Beghi A. Machine learning for predictive maintenance: A multiple classifier approach. IEEE Transactions on Industrial Informatics. 2015;11(3):812-820. doi:10.1109/TII.2014.2349359
15. Jardine AKS, Lin D, Banjevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing. 2006;20(7):1483-1510. doi:10.1016/j.ymssp.2005.09.012
16. Tasleem N, Raghav RS, Ansari MN, Sharma AJ. A decision intelligence framework: Integrating human intuition with AI models. Journal of Artificial Intelligence General Science. 2024;7(1):304-319. doi:10.60087/jaigs.v7i01.361
17. Lee J, Wu F, Zhao W, Ghaffari M, Liao L, Siegel D. Prognostics and health management design for rotary machinery systems—Reviews, 42(1–2), 314–334. 2014. doi:10.1016/j.ymssp.2013.06.004
18. Tasleem N, Ansari MN, Afzal MNI. AI-driven decision intelligence: A Bayesian framework for valuing HR technology readiness and international investment under uncertainty. Journal of Modelling in Management. 2026:1-20. doi:10.1108/JM2-11-2025-0604
19. Carvalho TP, Soares FAAMN, Vita R, Francisco RP, Basto JP, Alcalá SGS. A systematic literature review of

machine learning methods applied to predictive maintenance. Computers & Industrial Engineering. 2019;137:106024. doi:10.1016/j.cie.2019.106024
20. Qin SJ. Survey on data-driven industrial process monitoring and diagnosis. Annual Reviews in Control. 2012;36(2):220-234. doi:10.1016/j.arcontrol.2012.09.004
21. Venkatasubramanian V, Rengaswamy R, Yin K, Kavuri SN. A review of process fault detection and diagnosis: Part I—Quantitative model-based methods. Computers & Chemical Engineering. 2003;27(3):293-311. doi:10.1016/S0098-1354(02)00160-6
22. Venkatasubramanian V, Rengaswamy R, Kavuri SN. A review of process fault detection and diagnosis: Part II—Qualitative models and search strategies. Computers & Chemical Engineering. 2003;27(3):313-326. doi:10.1016/S0098-1354(02)00161-8
23. Venkatasubramanian V, Rengaswamy R, Kavuri SN, Yin K. A review of process fault detection and diagnosis: Part III—Process history based methods. Computers & Chemical Engineering. 2003;27(3):327-346. doi:10.1016/S0098-1354(02)00162-X

How to Cite This Article
Hande S, Waman S, Bankar S, Haral S, Khatal SS. Industrial knowledge graph framework for autonomous root cause analysis and predictive process optimization. Int J Future Eng Innov. 2026;3(6):76-81.

Creative Commons (CC) License
This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms